مدخل لدراسة و تحليل البيانات الإحصائية

تطبيقات على برنامج

SPSS

جمال شعوان

2014

الكتاب : مدخل لدراسة و تحليل البيانات الإحصائية.
تطبيقات على برنامج SPSS
تأليف : جمال شعوان Jamal CHAAOUAN
الإصدار الثاني : 2014
رخصة النشر : رخصة المشاع الإبداعي creative commons
موقع المؤلف : http://goo.gl/xci10
مدونة المؤلف : http://chaaouan.blogspot.com



أساليب التحليل الإحصائي للبيانات

1 – التحليل أحادي المتغيرة Analyse univarie

يعد التحليل أحادي المتغيرة من أسهل أنواع الأساليب الإحصائية، حيث يهتم بمتغير واحد فقط، حتى و إن كان عدد المتغيرات كبيرا، فإن هذا النوع من التحليل يدرس كل متغيرة على حدى لمراقبة جودة المعلومات من حيث خلوها من الأخطاء و التعرف على القيم الشاذة في البيانات و تلخيص المعلومات في شكل أرقام أو رسوم بيانية. كما يسمح لنا هذا النوع من التحليل من مقارنة النتائج المحصل عليها بين المتغيرات و تختلف أنواع التحليل باختلاف مقاييس المتغيرات من حيث كونها بيانات إسمية أو ترتيبية أو كمية.

1-1 - مقاييس النزعة المركزية

1-1-1 – المتوسط الحسابي La moyenne arithmétique

و يمثل مركز الثقل في أي البيانات الجغرافية، و الغرض من استخدامه هو الاستغناء عن استقراء مفردات المجموعة كلها، و يحسب بإيجاد مجموع قيم المفردات أو الحالات و قسمته على عددها (عدد المفردات). يعرّف المتوسط الحسابي على أنه مجموع قيم المفردات الخاصة بالمتغير في إطار العينة (x1, x2, x3, …. xn) مقسوم على عدد المفردات n (الحالات) و يرمز له بالرمز X و يحسب بتطبيق المعادلة التالية :

$$\overline{X} = \frac{\sum x_i}{\mathbf{n}}$$

x_i = المفردة أو الحالة

n = عدد المفردات

سنأخذ بيانات الجدول التالي كمثال لحساب المتوسط الحسابي بالمتسلسلات الإحصائية البسيطة :

التساقطات المطرية بالميلمتر بعشر محطات مناخية من المغرب

المحطة المناخية	التساقطات (ملم)
الحسيمة	380
الدار البيضاء	445
أكادير	230
الصويرة	222
فاس	523
مراكش	172
الرباط	560
وجدة	226
الناظور	443
الجديدة	409
n= 10	$\sum x_i$ **= 3610**

تتكون هذه المتسلسلة الإحصائية من 10 حالات cases، وتمثل "التساقطات السنوية" المتغيرة المدروسة بهذه العينة، و من الملاحظ أن التساقطات تتراوح ما بين 172 ملم كأدنى معدل و 560 كمعدل أعلى.

إذا أردنا حساب متوسط التساقطات بالمحطات المناخية العشر فسنقوم بقسمة مجموع كمية التساقطات بالمحطات على عشرة

أي

$$\overline{X} = \frac{\sum x_i}{\text{n}} = \frac{3610}{10} = 361$$

إذن متوسط التساقطات بالمجال المدروس هو 361 ملم

أما في حالة المتسلسلات الإحصائية المبوبة (بيانات عبارة عن فئات) حيث كل فئة يقابلها عدد من التكرارات فإن المتوسط الحسابي يحسب كالتالي:

$$\overline{X} = \frac{\sum F_i x_i}{\sum F_i}$$

x_i = مركز الفئة

F_i = تكرار الفئة

$\sum F_i$ أو n = التكرار الكلي

سنأخذ بيانات الجدول التالي كمثال لتوضيح كيفية التعامل مع الفئات و كيفية حساب مركز الفئة

توزيع مساحة الاستغلاليات الزراعية بالهكتار بالنسبة ل 281 استغلالية

حجم الاستغلالية بالهكتار	عدد الاستغلاليات الزراعية	x_i	$F_i x_i$
أقل من 4	14	2	28
[4-8[	61	6	366
[8-12[	94	10	940
[12-18[	90	15	1350
أكثر من 18	22	21	462
	$\sum F_i$ = 281		$\sum F_i x_i$ =3146

- حساب مركز الفئة x_i

مركز الفئة هو المتوسط الحسابي لطرفي الفئة. فمثلا مركز الفئة [4-8[يساوي مجموع الطرفين مقسوم على 2

$$x_{i[4-8[} = \frac{4+8}{2} = 6$$

لكن الأمر يختلف بالفئات المشار إليها بـــ: "أقل من" أو "أكثر من".

بالنسبة للفئة التي تتضمن عبارة "أقل من" فإن مركز الفئة يساوي الحد الأقصى للفئة ناقص 1/2 مدى الفئة التي بعدها.

ففي الفئة : أقل من 4 لدينا:

4 هي الحد الأقصى للفئة

و مدى الفئة التي بعدها هو مدى الفئة]4-8[، يساوي 4

إذن : $x_{i\,4\,\text{أقل من}} = 4 - \frac{4}{2} = 4 - 2 = 2$

أما بالنسبة للفئة "أكثر من" (فإنها عكس السابقة) فإن مركز الفئة يساوي الحد الأدنى للفئة زائد 1/2مدى الفئة التي قبلها.

ففي الفئة "أكثر من 18" لدينا

18 هو الحد الادنى للفئة

و مدى الفئة التي قبلها هو مدى الفئة]12-18[، يساوي 6

إذن

$$x_{i\,18\,\text{أكثر من}} = 18 + \frac{6}{2} = 18 + 3 = 21$$

فبعد حساب مركز الفئة x_i وضربه في التكرار المرافق له F_i يمكننا حساب المتوسط الحسابي

$$\overline{X} = \frac{\sum F_i x_i}{\sum F_i} = \frac{3146}{281} = 11,2$$

و عليه يمكن القول بان متوسط الاستغلاليات الزراعية هو في حدود 11,2 هكتار للاستغلالية الواحدة.

أما إذا أردنا حساب المتوسط الحسابي بعدد الاستغلاليات بكل فئة فإننا سنحسبه من خلال معادلة المتسلسلات البسيطة

$$\overline{X} = \frac{\sum F_i}{n} = \frac{281}{5} = 56,2$$

و سنرى لاحقا كيفية التعامل مع المتسلسلات المبوبة باستخدام برنامج التحليل الإحصائي SPSS

2-1-1 – الوسيط La médiane

هي القيمة التي تتوسط توزيع القيم في البيانات، حيث تقسم العينة إلى جزأين متساويين، و ذلك بعد ترتيب عناصر المجتمع إما تنازليا او تصاعديا، و يرمز له بالرمز Me و تختلف طرق حساب الوسيط حسب عدد القيم و نوعها.

- في الحالة التي يكون فيها مجموع المفردات عددا فرديا فإن حساب الوسيط يتم بإضافة رقم واحد لمجموع المفردات و قسمة المجموع الكلي على 2

$$\frac{n+1}{2}$$

حيث أن n تمثل عدد أفراد العينة .

سنأخذ قيم الجدول **1** كمثال و سنحذف منه قيمة محطة مراكش و ستصبح المتسلسلة الغحصائية على الشكل التالي بعد ترتيبها تصاعديا

222; 226; 230; 380; 409; 443; 445; 523; 560

بحيث n=9 و بالتالي يصبح الوسيط

$$\frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5$$

أي أن قيمة الوسيط هي قيمة n_5 و التي تساوي **409**

- في الحالة التي يكون فيها مجموع عدد أفراد العينة عددا زوجيا فإن قيمة الوسيط هي المتوسط الحسابي للمفردتين اللتين ترتيبهما $\frac{n}{2}$ و $(\frac{n}{2}+1)$

أي :
$$Me = \frac{\frac{n}{2}+(\frac{n}{2}+1)}{2}$$

نأخذ كمثال قيم المتسلسلة الإحصائية بالجدول الأول و نرتبها تصاعديا فتصبح كالتالي :

172; 222; 226; 230; 380; 409; 443; 445; 523; 560

سنجد بأن n=10 أي أن n عدد زوجي و بالتالي فإن قيمة الوسيط تساوي

$$Me = \frac{\frac{10}{2}+(\frac{10}{2}+1)}{2} = \frac{380+409}{2} = 394,5$$

- لكن في المتسلسلات المبوبة (الممثلة بالفئات) فإن الوسيط يتم حسابه من خلال المعادلة التالية :

$$Me = L_U + \frac{\sum \frac{F_i}{2} - F}{fm} * i$$

Lo : تمثل الحد الأدنى للفئة الوسيطة

F : التكرار المتجمع الصاعد للفئة قبل الفئة الوسيطة

fm : تكرار الفئة الوسيطة

I : مدى الفئة الوسيطة

نأخذ الجدول التالي كمثال و نحسب فيه قيمة الوسيط.

حجم الاستغلالية بالهكتار	عدد الاستغلاليات الزراعية	مركز الفئة x_i	التكرار الصاعد	التكرار النسبي المتجمع الصاعد	التكرار النسبي المتجمع النازل
أقل من 4	14	2	14	4,98	**100**
[4-8[	61	6	75	26,69	95,02
[8-12[	94	10	169	60,14	73,31
[12-18[	90	15	259	92,17	39,86
أكثر من 18	22	21	281	**100**	7,83
	$\sum F_i = 281$				

لدينا n = Fi = 281

إذن

$$\frac{n+1}{2} = \frac{281+1}{2} = 141$$

و من خلال التكرار الصاعد لدينا **169** هي القيمة الأقرب ل **141** و بالتالي فإن الفئة]8-12] هي الفئة الوسيطة، و عليه نحسب قيمة الوسيط بتطبيق المعادلة المناسبة لذلك

أي

$$Me = L_0 + \frac{\sum \frac{F_i+1}{2} - F}{fm} * i = 8 + \frac{141-75}{94} * (12-8) = 10{,}81$$

إذن قيمة الوسيط تساوي 10,81 و هو أقل نسبيا من الوسيط (11,2) و يمكن حساب هذه القيمة من خلال الرسم البياني بتمثيل التكرارات النسبية المتجمعة الصاعدة و التكرارات النسبية المتجمعة النازلة و تأخذ قيمة الوسيط من خلال إحداثيات نقطة تقاطع منحنيي التكرار.

3-1-1 – المنوال Le Mode

هو القيمة التي لها أكبر تكرار في المتسلسلة الإحصائية و قد يكون للمتسلسلة الإحصائية أكثر من منوال. أما في البيانات المبوبة فإن قيمته تحسب بالمعادلة التالية :

$$M = L_0 + \frac{D_1}{D_1 + D_2} * i$$

Lo : تمثل الحد الأدنى للفئة المنوالية

D1 : الفرق بين تكرار الفئة المنوالية و الفئة التي قبلها

D2 : الفرق بين تكرار الفئة المنوالية و الفئة التي بعدها

i : مدى الفئة المنوالية

في المتسلسلات البسيطة تمثل المنوال القيمة الأكثر تواترا بعد ترتيب المتسلسلات ترتيبا تصاعديا أو تنازليا.

2-1 – مقاييس التشتت Paramètres de dispersion

تبحث مقاييس التشتت في كيفية التعرف على مقدار انتشار البيانات أو تبعثرها، فهي على عكس مقاييس النزعة المركزية التي تتمحور حول قيم مركزية بين مجموعة من المتغيرات، فمقاييس النزعة المركزية وحدها لا تكفي لتقديم فكرة دقيقة عن توزيع البيانات بالمجتمع الإحصائي، و لهذا نلجأ إلى دراسة التوزيع باعتماد مقاييس التشتت و الانتشار للتعرف على مدى تشتت مفردات المتسلسلة الإحصائية حول وسطها الحسابي في العينة، و كلما اقتربت قيم المقاييس من الصفر كلما كان التشتت ضعيفا، و العكس صحيح.
أهم هذه المقاييس الخاصة بدراسة التشتت هي التباين و الانحراف المعياري و معامل التشتت و المدى.

1-2-1 – المدى Range

يعتبر من أبسط مقاييس التشتت إذ يعطينا فكرة عن المدى الذي يمكن أن تتشتت به قيم المتسلسلة الإحصائية، و يقترن بأدنى قيمة و أعلى قيمة، و يرمز له بالرمز E و يتم حسابه بطرح أصغر قيمة في العينة من أكبر قيمة في نفس العينة.

xi - E = xn

xn أكبر قيمة في العينة

xi أصغر قيمة في العينة

و توافق قيم xn و xi على التوالي في المتسلسلات المبوبة مركز الفئة الأخيرة و مركز الفئة الأولى

2-2-1 – التباين La Variance

يعرف بأنه المتوسط الحسابي لمربعات الانحرافات عن المتوسط الحسابي، و يصطلح عليه في بعض الأحيان "بتباين المجتمع" و يرمز له بالرمز Var(x) أو $(\sigma x)^2$ و يحسب من خلال المعادلة التالية :

في حالة البيانات البسيطة

$$\mathrm{Var}(x) = \frac{\sum(x_i - \bar{x})^2}{n}$$

في حالة المتسلسلات المبوبة :

$$\mathrm{Var}(x) = \frac{\sum Fi(x_i - \bar{x})^2}{\sum F_i}$$

3-2-1 – الانحراف المعياري Ecart-type

يعطي فكرة عن تشتت القيم عن متوسطها الحسابي، و يعادل الجذر التربيعي للتباين و يصطلح عليه كذلك بتباين العينة، أي أن تباين العينة ecart-type يساوي الجذر التربيعي لتباين المجتمع variance

$$\sigma x = \sqrt{Var(x)} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

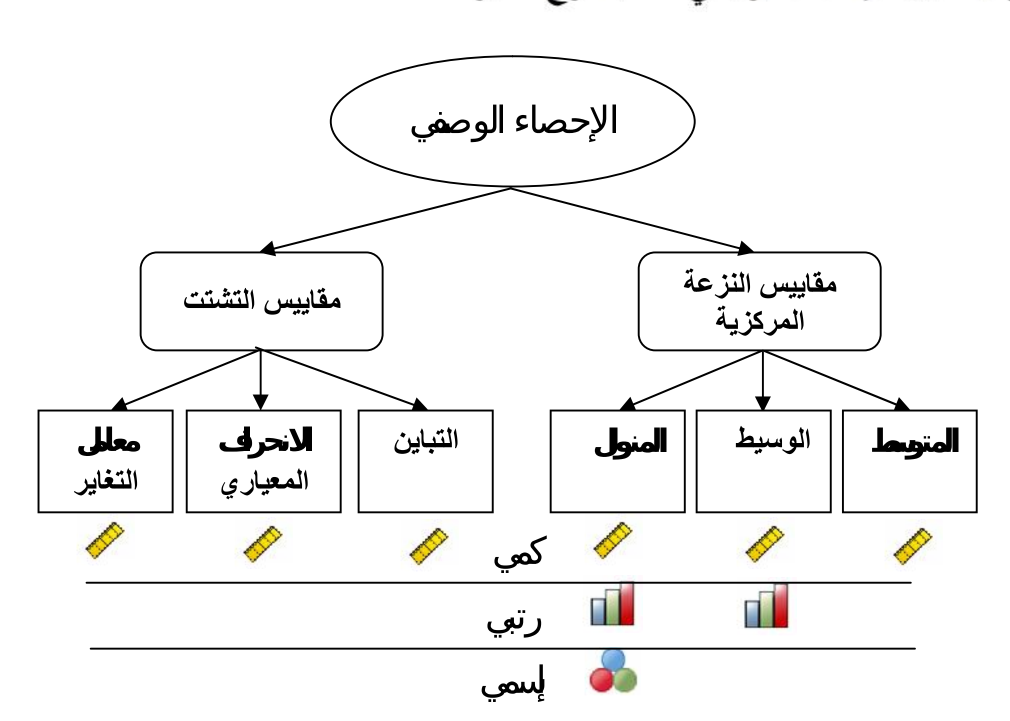
4-2-1 – معامل التشتت Coefficient de variation

يسمى كذلك بمعامل التغير، و يستخدم لتوضيح نسبة تشتت القيم بالعينة المدروسة، و يتم حسابه من خلال المتوسط الحسابي و الانحراف المعياري، إذ يتم قسمة الانحراف المعياري على المتوسط الحسابي و ضربه في **100** للحصول على النسبة المئوية

$$CV = \frac{\sigma x}{\bar{x}} * 100$$

3-1 - أساليب التحليل بمتغير واحد حسب نوع البيانات

الإلمام الجيد بالدراسة يسهل على الباحث تحديد نوع التحليل الذي سيستخدمه، و في كل مستوى من مستويات التحليل يعد معرفة نوع المتغيرات ضروريا لمعرفة نوع المقاييس و الأساليب الإحصائية التي تناسب كل متغيرة، و خاصة عند استخدام البرامج الإحصائية، إذ أنها تسمح بإنجاز قياسات مغلوطة، كحساب المتوسط الحسابي للمتغيرات النوعية، لأنه لطالما يتم ترميز متغير كالجنس مثلا من أجل إخضاعه لعمليات التحليل فإن البرنامج سينفذ جل العمليات على أساس الأرقام التي تم بها الترميز. و فيما يلي تلخيص لأساليب الإحصاء الوصفي حسب نوع المتغيرة.



2 ـ أساليب التحليل ثنائي المتغيرة

على عكس المستوى الأول المتمثل في التحليل أحادي المتغيرة، الذي يدرس كل متغيرة على حدى، فإن المستوى الثاني يندرج ضمنه التحليل الثنائي الذي ينبني على دراسة العلاقات الثنائية بين متغيرين فقط، حتى و إن فاق عدد المتغيرات 2 فإنه يهتم بدراسة كل متغيرتين على حدى. اهم أنواع التحليل بهذا المستوى و هو الارتباط. و تختلف أساليب التحليل ثنائي المتغيرة حسب نوع البيانات من حيث كونها إسمية أو ترتيبية أو كمية.

1-2 – الارتباط La corrélation

يهتم الارتباط بدراسة العلاقة بين متغيرتين، بحيث إذا تغير أحدهما مال الآخر إلى التغير.

- في الحالة بالارتباط الطردي (ارتباط الموجب) فإن العلاقة بين المتغيرتين هي علاقة طردية موجبة، كلما زادت قيمة المتغيرة الأولى (المتغيرة المستقلة) زادت بالمقابل قيمة المتغيرة الثانية (المغيرة التابعة).
- أما إذا كان التغير في الاتجاه المعاكس، فيسمى بالارتباط العكسي أو ارتباط سالب، اي أن العلاقة بين المتغيرتين هي علاقة عكسية أو سالبة، حيث كلما زادت قيمة المتغير الأول قلت قيمة المتغير التابع أو العكس، كلما قلت قيمة المتغير الأول زادت قيمة المتغير التابع.

يعد الارتباط نوعا من المقاييس الإحصائية الأكثر شيوعا في الدراسات الجغرافية، و خصوصا الارتباط الخطي البسيط لبيرسون Pearson الذي يدرس العلاقة بين متغيرتين كميتين، فكلما اقتربت قيمته من 1 دل ذلك على وجود علاقة ارتباط قوية، و كلما اقتربت من الصفر دلّ ذلك على ضعف أو انعدام العلاقة بين المتغيرتين، فيما أن إشارة معامل الارتباط تدل على نوع العلاقة طردية أو عكسية (موجبة أو سالبة). و يمكن حساب قوة العلاقة بين متغيرتين كميتين x و y و قياس الارتباط بينهما من خلال المعادلة التالية:

$$r = \frac{\text{cov}(x,y)}{\sigma x * \sigma y}$$

r : معامل الارتباط

σx : الانحراف المعياري للمتغيرة x

σy : الانحراف المعياري للمتغيرة y

Cov(x,y) : التباين المشترك covariance ل x و y و يتم حسابه باستخدام إحدى المعادلات التالية :

$$cov(x,y) = \frac{\sum(x_i - \bar{x})(y_j - \bar{y})}{n}$$

أو

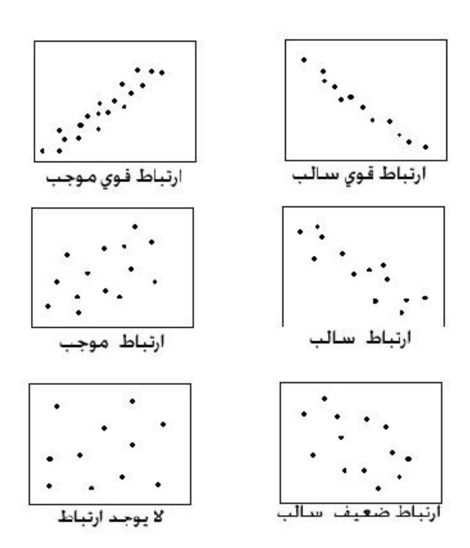
$$cov(x,y) = \left(\frac{\sum(x_i * y_j)}{n}\right) - (\bar{x} * \bar{y})$$

أما في حالة المتسلسلات المبوبة نستخدم المعادلة التالية :

$$cov(x,y) = \frac{\sum F_i(x_i - \bar{x})(y_j - \bar{y})}{\sum F_i}$$

يشترط في الارتباط استقلالية أفراد العينة، و في حالة عدم اعتدال المتغيرتين نستخدم معامل ارتباط آخر (سبيرمان و كندال تاو). ويتم حساب معامل الارتباط البسيط لسبيرمان r من خلال قسمة التباين المشترك ل x و y على حاصل ضرب الانحراف المعياري لكل من x و y

تكون قيمة الارتباط محصورة بين 1+ و 1- و تأخذ العلاقة بين المتغيرتين الأشكال التالية حسب إشارة معامل الارتباط.



2-2 معامل ارتباط الرتب (Spearman)

يعرف بمعامل ارتباط سبيرمان Spearman للرتب، و لذا تختلف قيمته عن قيمة معامل بيرسون (للقيم الأصلية و ليس لرتبها) و هو أقل دقة من معامل ارتباط بيرسون و يتعامل مع البيانات الرقمية و غير الرقمية للترتيب مثل : ضعيف ، متوسط ، جيد .. و يرمز له بالرمز rs و يدخل ضمن الإحصاءات غير المعلمية Non-paramétrique ذات التوزيع الحر و قيمته موجبة أقل أو تساوي الواحد الصحيح، و تحسب قيمته باستخدام المعادلة التالية :

$$r_s = 1 - \frac{\sigma - \sum D^2}{n(n^2 - 1)}$$

n : عدد المشاهدات

D : الفرق بين رتبتي كل قيمتين متقابلتين

و يعتمد معامل الارتباط الرتبي على رتب مستويات المتغيرين للقيم الأصلية من خلال ترتيب مفردات كل متغير من المتغيرات الترتيبية ترتيبا تصاعديا أو تنازليا مع إعطاء كل مفردة قيمة تبين ترتيبها.

3-2 تحليل كاي تربيع X^2

تقوم فكرة مربع كاي على أساس مقارنة البيانات، أي المشاهدات الفعلية التي تمت مشاهدتها من طرف الباحث ببيانات أخرى متوقعة و التي تعبر عن الفرضيات التي وضعها الباحث، فإذا كانت قيمة (x^2) المحسوبة كبيرة، فإن الفرضية الموضوعة غير صحيحة، أما إذا كانت قيمة x^2 صغيرة فإن الفرضية تكون صحيحة لأن الفروق بين التكرارات المشاهدة و المتوقعة تكون قليلة.

و يستخدم كاي تربيع x^2 (مربع كاي) للبيانات المعبر عنها بالتكرار في مستويين أو أكثر كالذكور و الإناث أو مدينة و قرية،.. و هدفه حساب معامل الارتباط لنسبتين أو أكثر لمتغير واحد (أحادي) أو متغيرين تصنيفيين يضم كل منهما مستويين أو أكثر (ثنائي).

و معادلته كاتالي :

$$x^2 = \frac{\Sigma(Q_i - E_i)^2}{E_i}$$

حيث أن

Qi : التكرار المشاهد الذي نحصل عليه من خلال الجرد الميداني

Ei: التكرار المتوقع = $\frac{\text{مجموع التكرارات الأفقية} \times \text{مجموع التكرارات العمودية}}{\text{المجموع الكلي للتكرارات}}$

n عدد درجات الحرية = (عدد الأسطر - **1**) (عدد الأعمدة – **1**)

أما قيمة كاي تربيع الجدولية تستخرج من جدول x^2 حسب درجة الحرية n و مجال الثقة α الذي يحدده الباحث و غالبا ما يكون 0,5 أو 0,1

4-2 – معامل فاي Phi Corrélation

نستخدمه في حالة دراسة الارتباط بين متغيرات غير قابلة للترتيب كمتغير الجنس أو التدخين ... و يمكن حساب معامل فاي انطلاقا من مربع كاي و عدد الافراد.

$$r_o = \sqrt{\frac{x^2}{n}}$$

5-2 – معامل التوافق
إذا كان لأحد المغيرتين على الأقل أكثر من صفتين (كاللون مثلا) عندها يعرف معامل الاقتران بمعامل التوافق، و يرمز له بالرمز rc و يقاس الارتباط من الصيغة التالية التي تعتمد على حساب معامل x^2

$$r_c = \sqrt{\frac{x^2}{x^2 + n}}$$

3 – التحليل متعدد المتغيرات Analyse multivarié
أدى قصور التحليل الاحادي و الثنائي المتغيرة في التعامل مع عدد كبير من المتغيرات إلى ظهور المستوى الثالث من التحليل، و هو التحليل متعدد المتغيرات، الذي يهتم بدراسة متغيرات متعددة أو مجموعة من المتغيرات في وقت واحد. و لعل التحليل العاملي Analyse factorielle من أهم الأساليب المستخدمة في هذا المستوى من التحليل، و يستخدم في تفسير العلاقات و تبسيط الارتباطات بين مختلف المتغيرات، حيث تقوم فكرة التحليل العاملي على تلخيص العديد من المتغيرات لعدد أقل يعرف بالعوامل les facteurs . هذا بالإضافة إلى نوع آخر من تحليل التباين، كتحليل فيشر للتباين ANOVA أو ما يسمى بتحليل التباين.

إذن فالتحليل العاملي هو أسلوب إحصائي يستهدف تفسير معاملات الارتباط التي لها دلالة إحصائية بين مختلف المتغيرات، وصولا إلى العوامل المشتركة التي تصف العلاقة بين المتغيرات و تفسيرها، حيث يهدف إلى تقليل حجم البيانات و تلخيصها و الإقلال من المتغيرات إلى عدد ضئيل من العوامل مستندا في ذلك إلى معامل الارتباط بين كل متغير و غيره من المتغيرات الأخرى على عكس التحليل ثنائي المتغيرة الذي يدرس الارتباط بين متغيرتين فقط.

II - الاستمارة Questionnaire
الاستمارة هي قائمة من الأسئلة المتنوعة تجيب عنها عينة من الناس ذوي الصلة بموضوع البحث للحصول على حقائق أو معلومات لأجل إنجاز بحث معين. و يجب أن تكون أسئلة الاستمارة بسيطة و مفهومة و مركزة، كما يجب الابتعاد عن الأسئلة الموحية للإجابة (أليس ... ؟، ألا تعتقد.....؟ ...) التي تدفع المجيب إلى الإجابة بنعم و كأن الباحث يريد تلك الإجابة بالضبط و هذا يدخل ضمن التأثير في الشخص المجيب لاختيار الجواب الذي يرغبه الباحث، و تحديد وحدات القياس عندما تكون الأجوبة عبارة عن أرقام (لتر، متر، درهم دولار، طن ...). و من الضروري أن تكون الأسئلة مباشرة وواضحة لا تتطلب من المجيب التفكير بعمق حتى يجيب على السؤال و حتى لا يصاب بالملل، و من الأفضل ان تكون

الاستمارة قصيرة قدر الإمكان لأنه قد لا يكون عند المجيب وقتا طويلا للإجابة على كافة الأسئلة، و يفضل أن توزع الاستمارة على مجموعة صغيرة للتجريب و تعديل الأخطاء قبل التطبيق النهائي.

يجب صياغة استمارة صادقة و ثابتة حتى يتسنى تعميم و استغلال البيانات لفترة زمنية معقولة. فبرامج التحليل الإحصائي تسمح لنا اختبار درجة ثبات البيانات للتأكد من صحتها.

2.1 - ترميز الاستمارة

ينبغي تحويل أجوبة الاستمارة إلى أرقام أو حروف يسهل إدخالها إلى الحاسوب و تسمى هذه العملية بعملية ترميز الاستمارة. و حسب برنامج SPSS فإن الأشخاص الذين قاموا بالإجابة على أسئلة الاستمارة يطلق عليهم إسم الحالات Cases (الأفراد) و كل سؤال في الاستمارة هو بمثابة متغيرة Variable و تشكل الإجابات قيم المتغيرة Valeur de variable.

سنتطرق فيما يلي لبعض أنواع الأسئلة التي قد تطرح بالاستمارة و كيفية التعامل معها لترميزها على البرامج الإحصائية.

الجنس : **ذكر** ☐ **أنثى** ☐

السن : ....

عدد الإخوة : .... **الذكور** .... **الإناث** ....

المهنة : **الأب** .... **الأم** ....

هل ترغب بمتابعة الدراسة بالجامعة **نعم** ☐ **لا** ☐

أماكن الاتصال بالأنترنيت

مكان الاتصال	صعب	نادر	دائم
المنزل			
نادي الانترنيت			
مقهى			

هل توافق أن يكون تسجيل الطالب في الجامعة عبر الانتيرنت؟

موافق بشدة ☐ **موافق** ☐ محايد ☐ **معارض** ☐ **معارض بشدة** ☐

ما هي أهم الهوايات التي تمارسها ؟

القراءة ☐ الرياضة ☐ الصيد ☐ **أخرى** ☐

الدخل الشهري

أقل من 500		500 – 1000		1000 – 2000		أكثر من 2000	

الحل : **

1) سؤال يسمح باختيار إجابة واحدة فقط

الجنس : **ذكر** ☐ **أنثى** ☐

عندما يكون لدينا سؤال يسمح فيه باختيار إجابة واحدة فقط، فإنه يكفي صياغته في متغيرة واحدة، و في هذا المثال لدينا متغير "الجنس" مكون من "ذكور" و "إناث"، يمكن صياغته على برنامج SPSS بإضافة متغيرة جديدة بإسم "الجنس" و نقوم بترميز الذكور بالرقم 1 و الإناث بالرقم 2. نستخدم الأرقام في الترميز لأن البرنامج يتعامل مع البيانات الرقمية، فإذا قمنا بإدخال بيانات الجنس على هيئة "نص" فلن نستطيع إجراء جل التحليلات الإحصائية و التمثيلات البيانية، لكن يجب تحديد صفة المتغيرة (كمية أم رتبية أم إسمية)، إذ أن في الأصل الرقم 2 هو ضعف الرقم 1، ولكن إذا قمنا باختيار صيغة البيانات الاسمية فإن جميع الأرقام التي تم استخدامها في الترميز تعتبر ذات نفس القيمة و متساوية و ليست لها أي دلالة كمية، بمعنى أن 2 = 1.

2) سؤال عبارة عن متغيرة كمية.

السن : ......

تتطلب الإجابة على هذا السؤال تحديد عمر الشخص و في هذه الحالة تعتبر المتغيرة متغيرة كمية متصلة، و في بعض الحالات يمكن تحويلها إلى فئات عمرية.

3) سؤال يتضمن أكثر من متغير كمي

عدد الإخوة : .... **الذكور** .... **الإناث** ....

في هذا السؤال يجب معرفة عدد الإخوة الذكور و عدد الإخوة الإناث و يشكلان متغيرتين كميتين و في نفس السؤال مطلوب معرفة المجموع الذي يشكل متغير ثالث ضمن هذا السؤال و لربح الوقت يمكن حسابه من خلال جمع عدد الإخوة الذكور و الإناث و ذلك لربح الوقت في إدخال البيانات.

4) سؤال مفتوح (متغيرة نوعية (إسمية))

المهنة : **الأب** .... **الأم** ....

تهم متغيرة المهنة في هذه الحالة الأب و الأم و بالتالي لا يمكن جمعهم في خانة واحدة لكون مهنة الأب ليست هي مهنة الأم، و بالتالي تصبح لدينا مهنة الأب متغيرة، و مهنة الأم متغيرة ثانية، و بالتالي

يجب صياغة السؤال على برنامج SPSS في متغيرتين و يمكن حصر نوع المهن التي تتطلب الدراسة التأكد من توفرها أم لا و بالتالي صياغة السؤال على شكل جدول. و لترميز هذا السؤال سيتم تحديد بعض المهن و ترميزها بأرقام، ولاختيار أي المهن سيتم الاعتماد عليها فإنه يجب الأخذ بعين الاعتبار العوامل التالية :

- المهن الأكثر تكرارا في الاستمارة
- الهدف من الاستمارة
- شكل الاستمارة الذي تم تقديمه للأشخاص وكيفية الإجابة علية.
- الوقت المتاح للباحث.
- الدعم المادي المتوفر للباحث.
- الدقة المطلوبة.

5) سؤال منطقي يسمح باختيار إجابة واحدة فقط

هل ترغب بمتابعة الدراسة بالجامعة؟ **نعم** ☐ **لا** ☐

يعتبر هذا النوع من الأسئلة شبيه بالسؤال الأول إلا أن الامر يختلف في طبيعة المتغيرة، فإذا كان السؤال الأول متغيرة نوعية، فإن الأسئلة المنطقية التي تتطلب الإجابة بـــ "نعم" أو "لا" فغالبا ما يتم دراستها من الناحية الكمية و بالتالي ستختلف عن متغيرة الجنس في طبيعة القياس، و نستخدم قيمة **1** و **0** خلال ترميزها إذ تأخذ "نعم" القيمة "**1**" و "لا" تأخذ القيمة "**0**" و تأخذ وحدة القياس صيغة كمية Sclae

6) سؤال على شكل جدول

أماكن الاتصال بالأنترنيت

مكان الاتصال	**صعب**	**نادر**	**دائم**
المنزل			
نادي الانترنيت			
مقهى			

المطلوب في هذا السؤال تحديد درجة الاتصال بالانتيرنت من الأماكن الثلاث المذكورة و بالتالي سيتم التعامل مع هذا السؤال على أساس انه ثلاث أسئلة و ستحدد ثلاث متغيرات تتمثل في امكانية الاتصال بالانتيرنت من الأماكن الثلاث:

- الاتصال من المنزل
- الاتصال من نادي الانتيرنت
- الاتصال من المقهى

و كل متغيرة ستأخذ ثلاث قيم رتبية يجب ترميزها بأرقام يراعى فيها احترام التراتبية مثلا

- صعب = 1
- نادر = 2
- دائم = 3

و يجب تحديد المقياس الرتبي للمتغيرات.

7) سؤال رتبي

هل توافق أن يكون تسجيل الطالب في الجامعة عبر الانتيرنت؟

موافق بشدة ☐ **موافق** ☐ محايد ☐ **معارض** ☐ **معارض بشدة** ☐

هذا السؤال أشبه بالذي قبله فهو عبارة عن سؤال واحد يقبل جوابا واحدا، و يمكن ترميزه على شكل متغيرة واحدة (التسجيل عبر الانتيرنت) وذات أجوبة اختيارية رتبية ويمكن ان اعتماد.

- الرقم 5 ليدل على الإجابة " موافق بشدة"
- الرقم 4 ليدل على الإجابة " موافق"
- الرقم 3 ليدل على الإجابة " محايد"
- الرقم 2 ليدل على الإجابة " معارض"
- الرقم 1 ليدل على الإجابة " معارض بشدة".

8) سؤال يسمح بأكثر من إجابة

ما هي أهم الهوايات التي تمارسها ؟

القراءة ☐ الرياضة ☐ الصيد ☐ **أخرى** ☐

في هذا السؤال نلاحظ أن الشخص يمكن أن يجيب على أكثر من اختيار واحد، لذلك فان متغيرا واحدا لا يكفي لتمثيل السؤال. في هذه الحالة يفضل إنشاء خمسة متغيرات، كل متغير له احتمال إجابتين [نعم / لا] ويستخدم لهما:

- 1 للإجابة " نعم "
- و 0 للإجابة " لا" (أنظر السؤال 5)

9) سؤال عبارة عن فئات يقبل إجابة واحدة

الدخل الشهري

أقل من 500 ☐ 500 – 1000 ☐ 1000 – 2000 ☐ **أكثر من** 2000 ☐

المتغيرة المدروسة في هذه الحالة هي متغيرة الدخل الشهري و هي متغيرة كمية صيغت على شكل فئات، و لترميز الفئات نستخدم مركز الفئة.

مركز الفئة

مركز الفئة هو المتوسط الحسابي لطرفي الفئة. فمثلا مركز الفئة [500-1000[يساوي مجموع الطرفين مقسوم على 2

بالنسبة للفئة التي تتضمن عبارة "أقل من" فإن مركز الفئة يساوي الحد الأقصى للفئة ناقص 1/2 مدى الفئة التي بعدها. أما بالنسبة للفئة "أكثر من" (فإنها عكس السابقة) فإن مركز الفئة يساوي الحد الأدنى للفئة زائد 1/2مدى الفئة التي قبلها.

III - إدخال البيانات إلى برنامج SPSS

سنحاول في هذه الفقرة التعرف على كيفية إدخال البيانات إلى برنامج SPSS و كيف يمكن التعامل مع المتغيرات النوعية، لكون البرنامج يتعامل مع الأرقام في التحليل الإحصائي مما يتطلب منا ترميز البيانات حتى يتسنى لنا إخضاعها لعمليات التحليل.

لإدخال البيانات إلى برنامج SPSS أولا تحديد المتغيرات و تعريف طبيعتها للبرنامج بشكل صحيح لتفادي أخطاء المعالجة و التحليل، و يتم هذا العمل من خلال لوحة "التحكم في المتغيرات" " Affichage des variables" (Variable view) . تتكون هذه الوحة من سطور تمثل عدد المتغيرات، و أعمدة تحدد خصائص المتغيرة.

	1	2	3	4	5	6	7	8	9	10
	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure
1	الجنس	Numérique	8	0		{1, ذكر}...	Aucun	8	Droite	Nominales
2										Echelle
3										Ordinales
4										Nominales

1) إسم المتغيرة Nom

	Nom
1	الجنس
2	

يشترط في إسم المتغيرة أن يكون مقتضبا و أن لا يتعدى **68** حرفا و أن لا يبتدئ برقم و أن لا ينتهي بنقطة . أما بالنسبة للغة العربية فيجب ضبط البرنامج على UTF-8.

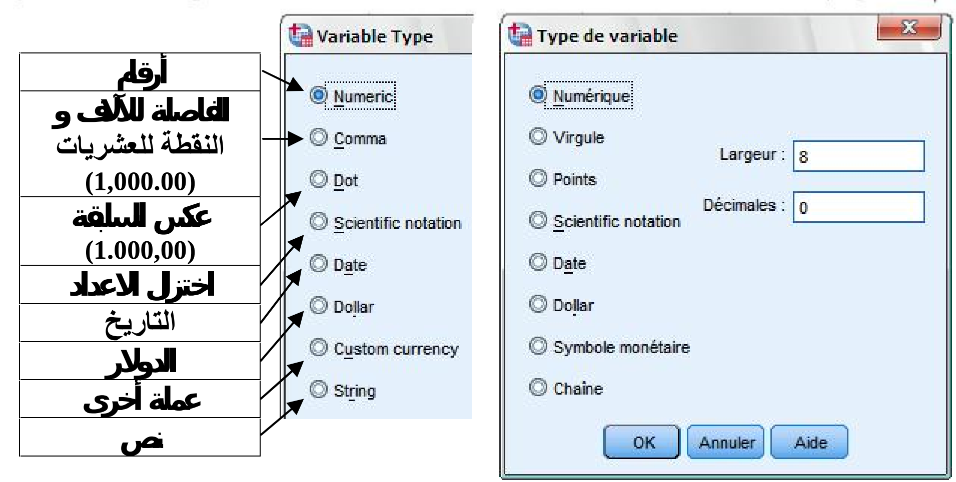
2) نوع المتغيرة Type

رغم تعدد الاختيارات التي يتيحها البرنامج إلا أنه في الأصل توجد ثلاث أنواع فقط و هي : الأرقام و النصوص و التاريخ. أما الباقي فهي عبارة عن تحسينات في كيفية إظهار البيانات.

الاختيار الأول و هو الاختيار الافتراضي في البرنامج لكونه أساس العمليات الاحصائية و تندرج ضمنه جميع أنواع المتغيرات المراد إخضاعها لعملية التحليل، حتي المتغيرات النوعية يتم إدخالها على شكل أرقام و يتم بعد ذلك ترميزها.

ليس هناك اختلاف في الجوهر فيما بين الأرقام و الفاصلة و النقطة فكلها تفي بنفس الغرض إلا أن الفرق بينهم يتجلى في كيفية عرض الأعداد، و هذا بطبيعة الحال لا يؤثر على عملية التحليل

Type
Numérique



لا نستعمل الاختيار الأخير إلا في حالة المتغيرات التي لا نريد أن نخضعها للتحليل. كأسماء الأفراد المستجوبين مثلا، لكون البرنامج لا يتعامل مع النصوص و الأسماء، و لهذا إذا أردنا إخضاع متغيرة نوعية للتحليل يجب ترميزها و اختيار الأمر الأول المتمثل في الأرقام Numérique.

3)طول النص او الأرقام (Largeur (Caractères

من خلال العمود Largeur الخاص بطول بيانات المتغيرة، يمكن حصر عدد الأرقام أو الحروف التي يمكن إدخالها و ذلك لتقليل الوقوع في الأخطاء أثناء عملية إدخال البيانات.

4) عدد الأرقام بعد الفاصلة Décimales

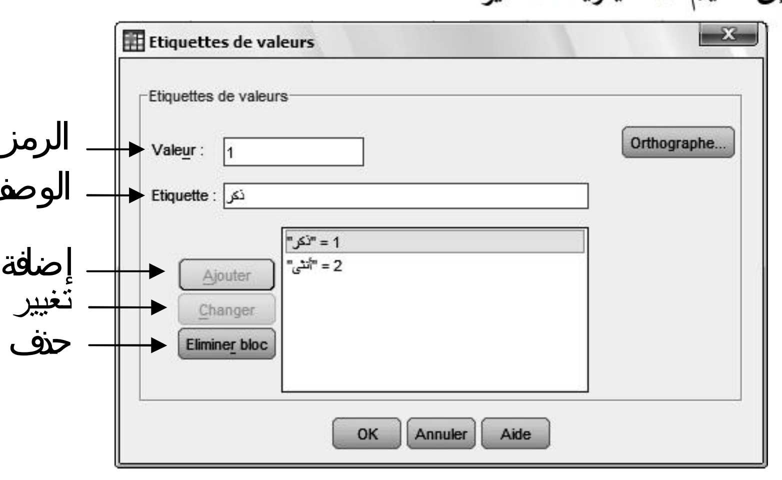
تهم خانة Décimales الأرقام فقط وذلك لتحديد عدد الأرقام بعد الفاصلة و يستحسن ضبطها على الصفر في حالة المتغيرات الكمية المنفصلة. كما أن البرنامج يقوم بعملية تصحيح الفاصلة في حالة البيانات المتصلة إذا لم يتم إظهار الأرقام العشرية.

5) الشارة Etiquette (*Label*)

تسمح لنا خانة Etiquette (*Label*) بكتابة الإسم الكامل للمتغيرة بشكل حر، لأن خانة الإسم مقيدة ببعض الشروط.

6) تعيين رموز للمتغير Valeurs

يستخدم هذا الأمر خلال مرحلة ترميز بيانات المتغيرة فنكتب في خانة Etiquette اسم الحالة، و في خانة Valeur نكتب الرمز الافتراضي الذي نود أن نعطيه لها. بعدها نضغط على Ajouter (Add) لإضافة القيمة الجديدة إلى القيم الاختيارية للمتغيرة.



7) القيم المفقودة *(Messing)* Manquant

أحيانا قد يقوم بعض الأشخاص بعدم الإجابة على سؤال ما فتبقى إجابة ذلك السؤال مفقودة وتسمى بالقيمة المفقودة، ويجب إبلاغ SPSS بذلك، وهناك عدة طرق لتعيين القيم المفقودة، نذكر منها:

عندما يكون هناك سؤال ليس له إجابة يجب تخطيه، ليقوم محرر البيانات بعرض تلك الخلية المفقودة بنقطة، وتسمى تلك القيم المفقودة " قيم نظام مفقودة (System Missing Values) " وجدير بالذكر انه بالنسبة للمتغيرات الرقمية فان الخلايا تحول إلى قيم نظام مفقودة ، أما بالنسبة للمتغيرات النصية فان الخلايا الفارغة تعامل كقيمة صحيحة، بمعنى آخر لا يوجد قيم مفقودة في المتغيرات النصية.

يمكنك أن تضع رمزا بدل القيم المفقودة لتصبح تلك القيم " قيم المستخدم المفقودة User Missing Values

Valeurs manquantes

Aucune valeur manquante

Valeurs manquantes discrètes

Plage plus une valeur manquante discrète facultative

Faible : Elevée :

Valeur discrète :

OK Annuler Aide

و يظهر من مربع الحوار عدة خيارات لتعيين القيم المفقودة كالتالي:

Aucune valeur manquante (No missing values)

يتم اختياره عند عدم وجود قيم مستخدم مفقودة وعادة يكون هذا الخيار محددا.

Valeurs manquantes discrètes (Discrete missing values)

يمكنك إدخال حتى ثلاث قيم مختلفة لمتغير واحد تعامل كقيم مستخدم مفقودة وهذا الخيار يصلح للمتغيرات الرقمية والنصية.

Plage plus une valeur manquante (Range of missing values)

يمكنك هذا الخيار من تحديد مدى معين من قيم المستخدم المفقودة بحيث تعامل اقل قيمة واكبر قيمة وما بينهما من القيم كقيم مفقودة. ويصلح هذا الخيار فقط للقيم الرقمية ولا يصلح للمتغيرات النصية. وجدير بالذكر أن قيم المستخدم المفقودة لا تدخل في الحسابات.

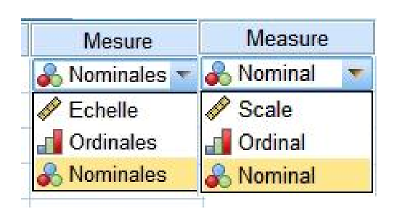
8) عرض العمود Colonnes *(Columns)*

و يتم تحديد عرض العمود و يقاس بعدد الأحرف التي يسمح لها بالظهور خلال مرحلة إدخال البيانات و لا يؤثر على البيانات.

9) موقع البيانات Align

موقع البيانات داخل العمود Alignement بحيث يمكن توجيهها لتكون في يسار العمود أو في وسطه أو في يمينه، و من الأفضل تركها على الحالة الافتراضية التي يختارها البرنامج تلقائيا، لأنه يقوم بمحاذاة النصوص إلى اليسار و الأعداد إلى اليمين، و هذا قد يساعدك في اكتشاف بعض الأخطاء المتمثلة في اختيار نوع المتغيرة.

10) المقياس Mesure (*Measure*)



لتحديد مقياس المتغير نضغط داخل الخلية اسفل Mesure ثم نضغط على السهم الموجود داخل الخلية فتظهر الخيارات التالية كما بالشكل.

المقياس الكمي Echelle (Scale) : يستخدم مع المتغيرات الكمية التي لم تخضع لعملية الترميز و هو الاختيار افتراضي من طرف البرنامج.

المقياس الرتبي Ordinales (Ordinal) : يتم اختياره في حالة المتغيرات النوعية المراد احترام التراتبية في حالاتها الإحصائية.

المقياس الإسمي Nominales (Nominal) : يجب اختياره عند ترميز المتغيرات الإسمية، كإشارة للبرنامج على أن القيم التي اعتمادها في الترميز هي متساوية.

بعد الانتهاء من كتابة المتغيرات ننتقل إلى مرحلة إدخال قيم المتخيرات في لوحة إظهار البيانات Affichage des données (Data View) و هي شبيهة إلى حد ما بلوحة برنامج Excel .

2 - إدخال بيانات المتغيرات النوعية.

بالنسبة للبيانات المرمزة الخاصة بالمتغيرات النوعية يمكننا إدخال البيانات بثلاث طرق :

- الطريقة الأولى : يتم كتابة الرمز الذي تأخذه كل صفة.
- الطريقة الثانية: يتم كتابة الصفة، ويشترط أن تكتب بشكل صحيح كما كتبت أثناء عملية الترميز، لأن أي خطأ سيقحمنا في القيم المفقودة.
- الطريقة الثالثة: اختيار الصفة من القائمة المنسدلة التي تظهر بعد النقر داخل الخانة المراد ملؤها. و تعتبرهذه الطريقة كحل بديل لتجاوز مشكلة إدخال بيانات غير معروفة.

تتطلب الطريقتين الثانية و الثالثة ضبط البرنامج على وضعية إظهار أسماء القيم المرمزة و ذلك من خلال قائمة Affichage (View) و اختيار الأمر Etiquettes de valeurs (Value Labels) أو بالضغط على الأيقونة الموافقة لهذا الأمر في شريط الأدوات التي توافقها الأيقونة التالية في النسخ القديمة.

1.2 - إدخال بيانات المتسلسلات الإحصائية المبوبة

نأخذ الجدول التالي كمثال

Y	X
320	[40-80[
1600	[80-120[
748	[120-160[
$\sum F_i$ = **2668**	

بالنسبة للجدول أعلاه لدينا متغيرتين X و Y، لو أردنا إدخالها إلى برنامج SPSS فإنه يجب علينا ترميز المتغيرة X بمركز الفئة و هي عبارة عن متغير رتبي Ordinale. بينما تشكل المتغيرة Y عدد الملاحظات أو التكرارات فيتم إدراجها كمتغير كمي Echelle.

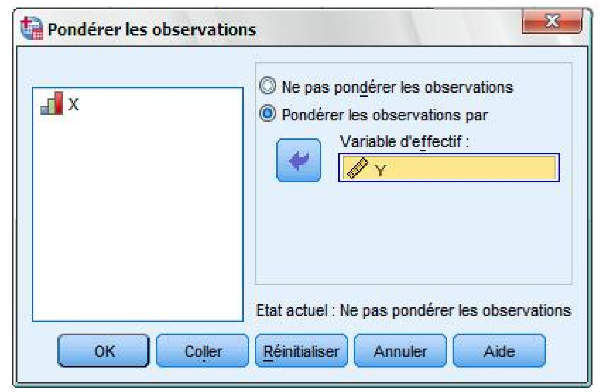
ففي هذه الحالة لإعطاء المتسلسلة الإحصائية وزنها الطبيعي ينبغي إدخال الفئة الأولى من المتغيرة X **320** مرة، و الفئة الثانية **1600** مرة، و الفئة الثالثة 748 مرة، بما مجموعه **2668** حالة Cases على أن كل حالة ستاخذ قيمة **1** كتكرار يقابلها في المتغيرة Y، و بطبيعة الحال سيتطلب هذا منا وقتا طويلا.

الحل البديل، هو أن يتم إدخال معطيات الجدول كما هي بعد ترميز المتغيرة X

5 : Y	X	Y
1	[40-80[	320
2	[80-120[	1600
3	[120-160[	748
4		

ثم من قائمة Données (Data) نختار الأمر Pondérer les observation (Weight Cases) أو الضغط على أيقونة

الميزان في شريط الأدوات أو و ذلك لإعطاء المتغيرة X وزنها من خلال التكرارات الموافقة لها بالمتغيرة Y . فتظهر لنا علبة الحوار التالية Boite de dialogue



فنختار المتغيرة التي تمثل التكرارات و في هذا المثال لدينا المتغيرة Y بعدها نضغط على Ok. لإلغاء أمر ترجيح الحالات نقوم بنفس الخطوات و نختار الأمر الأول Ne pas pondérer les observations (Do not weight cases)

IV - تطبيق بعض المقاييس الإحصائية باستخدام برنامج SPSS

بعدما تعرفنا على أنواع البيانات و كيفية إدخالها إلى برنامج SPSS سنحاول في هذه الفقرة التعرف على كيفية حساب مقاييس النزعة المركزية و مقاييس التشتت و الارتباط من خلال برنامج SPSS

1.4 - حساب مقاييس النزعة المركزية

سنأخذ الجدول التالي كمثال لحساب مقاييس النزعة المركزية

	المحطة	الارتفاع	الحرارة
1	أكنول	1210	1,5
2	أصيلا	12	7,0
3	باب برد	1220	,9
4	باب تازة	880	2,5
5	شفشاون	630	5,9
6	فاس	415	4,3
7	كتامة	1590	,5
8	العرائش	10	6,0
9	وزان	300	5,3
10	طنجة	75	9,0
11	تاونات	670	4,9
12	تازة	505	3,9
13	تطوان	5	8,3
14	زومي	350	4,4

يحتوي الجدول على 14 محطة مناخية بشمال المغرب تمتد على طول محور تازة العرائش و شمال هذا المحور، و يحتوي هذا الجدول على متغيرتين إحصائيتين (الارتفاع و درجة الحرارة)، فمقاييس النزعة المركزية تدخل ضمن التحليل أحادي المتغيرة أي أنه رغم توفر الجدول على أكثر من متغيرة فإننا سنقوم بدراسة مقاييس النزعة المركزية لكل متغيرة على حدى علما أننا سننفذ أمر التحليل جملة واحدة لكل التغيرات من خلال برنامج spss، و لا يعني ذلك أننا نقوم بالتحليل ثنائي المتغيرة بل التحليل أحادي المتغيرة الذي ينبني على دراسة و تفسير المتغيرات المتوفرة لدينا واحدة تلو الأخرى.

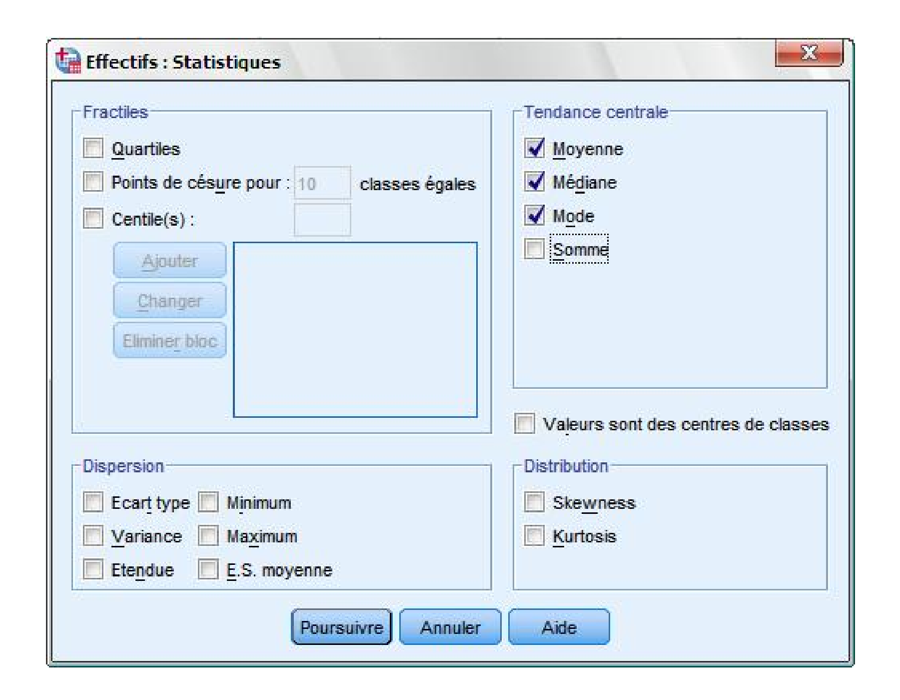
لحساب مقاييس النزعة المركزية باستخدام برنامج SPSS يتم استدعاء محلل البيانات، من شريط القوائم نختار الأمر Analyse (Analyze) ثم نختار مجموعة المقاييس الوصفية Statistique descriptive (Descriptive Statistics) و منها نختار الأمر Effectifs (Frequencies) فيظهر مربع حواري جديد مقسم إلى قسمين:

الأول على اليسار يظهر كل المتغيرات الواردة بالجدول التي يمكن إخضاعها للتحليل،

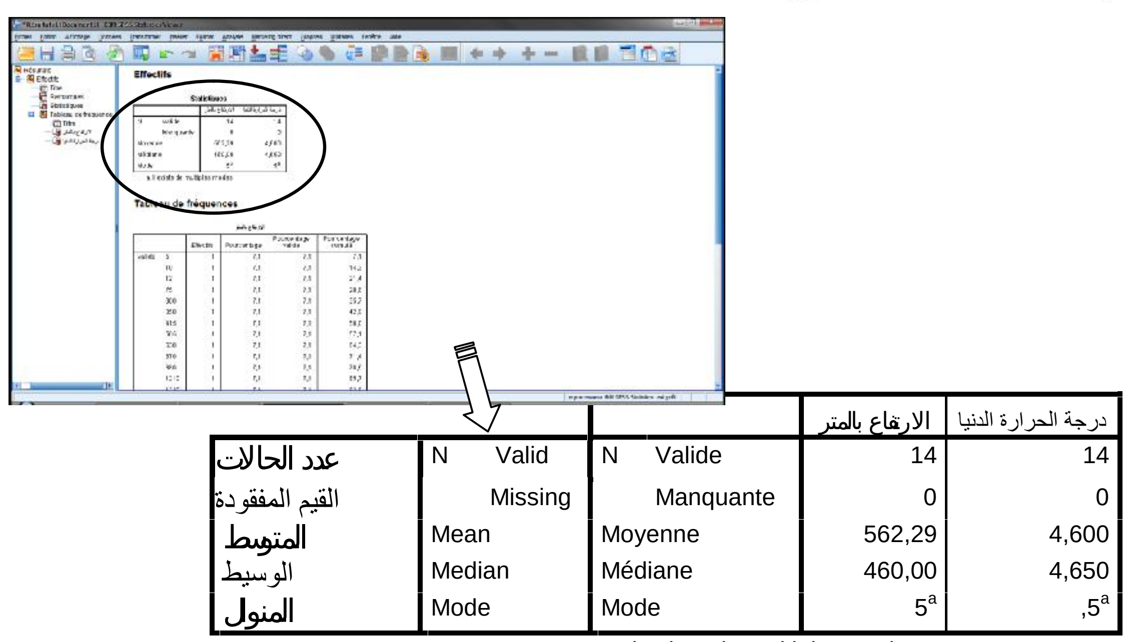
و الثاني يظهر المتغيرات التي تم انتقاؤها لإجراء العملية عليها و يتم ذلك من خلال انتقاء المتغيرة في العمود الأول و النقر عليها مرتين أو انتقاؤها و نقلها بالسهم الموجود بين القسمين.



بعد ذلك نختار الأمر Statistiques (Statistics) من نفس مربع الحوار فيظهر لنا مربع حواري آخر يظهر معظم مقاييس النزعة المركزية و التشتت.



سنختار في هذا المثال حساب المتوسط و الوسيط و المنوال وسنترك خانة المجموع فارغة لكون حساب مجموع الارتفاعات أو مجموع الحرارة في هذه الحالة يعد غير ضروري و ليس له أي دلالة إحصائية و بالتالي سنقتصر على المقاييس الثلاث الأولى. بعد اختيار المقاييس المراد حسابها نضغط على Poursuivre ثم Ok فتظهر لنا نافذة النتائج



			الارتفاع بالمتر	درجة الحرارة الدنيا
عدد الحالات	N Valid	N Valide	14	14
القيم المفقودة	Missing	Manquante	0	0
المتوسط	Mean	Moyenne	562,29	4,600
الوسيط	Median	Médiane	460,00	4,650
المنوال	Mode	Mode	5[a]	,5[a]

a. Il existe de multiples modes

يتضح من جدول النتائج أن عدد الحالات التي أخضعت للتحليل وصل إلى **14** حالة من دون وجود قيم مفقود أي أن جميع المحطات قد أخضعت للتحليل. و يتضح من الجدول نفسه أن متوسط

الارتفاع بالمجال المدروس هو 562,29 متر و يبعد عن القيمة الوسيطة بحوالي **102** متر إذ تمثل **460** متر القيمة الوسيطة بهذه المتسلسلة الإحصائية. في حين أن البرنامج أشار إلى القيمة **5** كقيمة منوالية مع إعطاء إحالة تدل على وجود اكثر من منوال و أنه اعتمد على أصغر قيمة في المتسلسلة و التي تمثل محطة تطوان التي تعلو عن سطح البحر ب **5** أمتار فقط. بينما متوسط الحرارة الدنيا هو بحدود 4,6° الذي يقترب من الوسيط ب 0,05° مما يدل على أن قيم درجات الحرارة الدنيا بالمجال المدروس تتوزع بشكل طبيعي و يمثل الوسيط القيمة 4,65° بينما هناك أكثر من منوال و يظهر البرنامج أصغر قيمة كقيمة منوالية و هي 0,5° المسجلة بمحطة كتامة.

2.4 - حساب مقاييس التشتت

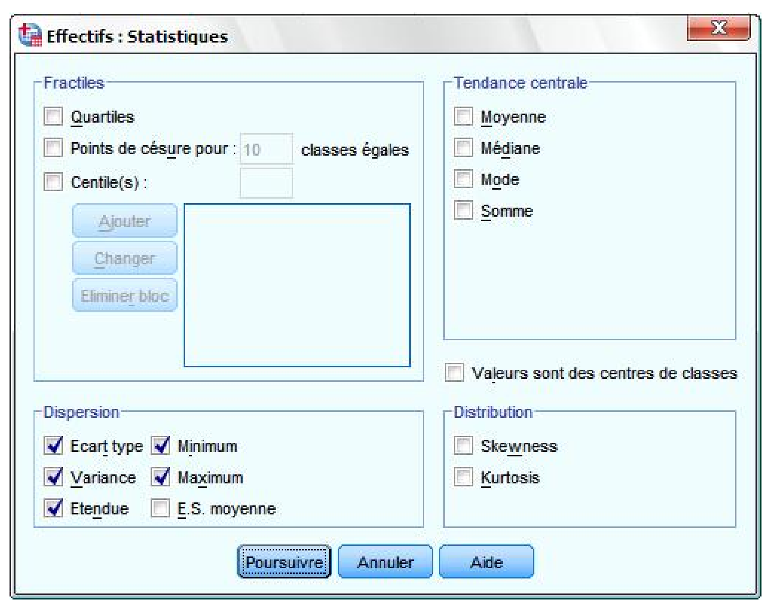
سبق و أن أشرنا إلى كيفية حساب مقاييس التشتت بالطريقة اليدوية، وسنحاول الآن معرفة كيفية حساب هذه المقاييس باستخدام برنامج SPSS.

بما أن هذه المقاييس تدخل ضمن مقاييس الإحصاء الوصفي فسنجدها على نفس المسار الذي قمنا من خلاله بحساب مقاييس النزعة المركزية. أي باتباع الخطوات التالية:

من قائمة Analyse نختار الأمر Statistiques descriptives ثم effectifs

يظهر لنا مربع حوار ونختار منه مربع الحوار التالية:

Equart-Type	*Standard deviation*	الانحراف المعياري
Variance	*Variance*	التباين (التغير)
Etendue	*Range*	المدى
Minimum	*Minimum*	القيمة الدنيا
Maximum	*Maximum*	القيمة القصوى



ثم نضغط على Poursuivre ثم Ok فتظهر لنا نافذة النتائج.

			الارتفاع بالمتر	درجة الحرارة الدنيا
عدد الحالات	*N Valid*	N Valide	14	14
القيم المفقودة	*Missing*	Manquante	0	0
الانحراف المعياري	*Std. Deviation*	Ecart-type	505,36	2,6044
التباين	*Variance*	Variance	255394,68	6,783
المدى	*Range*	Intervalle	1585	8,5
القيمة الدنيا	*Minimum*	Minimum	5	,5
القيمة القصوى	*Maximum*	Maximum	1590	9,0

نلاحظ من خلال الجدول التالي أنه لا توجد أية قيم مفقودة لكل من درجة الحرارة الدنيا و الارتفاع. بالنسبة للارتفاعات نلاحظ وجود مدى كبير يصل إلى **1585** متر إذ أن قيم الارتفاع بالمحطات المناخية تتراوح بين **5** أمتار و **1590** متر مما يدل أن المجال المدروس يمتد من بضعة أمتار عن سطح البحر حتى يصل إلى نطاق الغابة المختلطة من الأرز و السنديان. فوجود هذا الفارق الكبير في مدى الارتفاع قد يؤدي لا محالة إلى وجود اختلافات جذرية في درجة الحرارة الدنيا لاسيما و أن القاعدة تقضي بانخفاض درجة الحرارة مع الارتفاع بحوالي **1** درجة في كل **100** متر أو 0,5 في حالة وجود جو مشبع بالرطوبة، إلا أن هذه القاعدة قد تتأثر بالموقع العرضي أو البعد و القرب عن البحر و سنرى لاحقا إن أمكن دراسة الارتباط للتأكد من صحة وجود علاقة بين الارتفاع و درجة الحرارة.

تقوم مقاييس التشتت بدراسة تشتت قيم المتسلسلة الإحصائية عن المتوسط الحسابي، و لحساب نسبة تشتت القيم نقسم الانحراف المعيار على المتوسط الحسابي ثم نضربه في **100**

$$CV = \frac{\sigma x}{\bar{x}} * 100 = \frac{505,36}{562,29} * 100 = 89,9\%$$

و من خلال معامل التشتت تبين لنا أن قيم الارتفاع تعرف تشتتا كبيرا يصل إلى حوالي **90**%

3.4 - دراسة الارتباط الخطي البسيط لبيرسون Pearson

يمثل الجدول التالي مردودية الحبوب ببعض المناطق الفلاحية بالمغرب، و في المقابل تمت مقارنتها بمتوسط التساقطات، و نعلم أن التساقطات متغيرة مستقلة عن المردودية، بينما المردودية تعتبر متغيرة تابعة يمكن أن تتأثر بالتساقطات. و سندرس في هذا المثال مدى ارتباط أو مدى تأثر مردودية الحبوب بالتساقطات.

region	Larentabilite	P
Agadir	7	230
Alhoceima	18	380
Casa	20	445
Saouira	12	222
Fes	20	523
Marrakech	10	172
Ouejda	12	226
Rabat	20	560
Nador	17	443
Jadida	14	409

قمنا بدراسة الارتباط بين التساقطات و مردودية الحبوب بعشر مناطق مختلفة فكانت نتائج الارتباط كما يلي :

		La rentabilité des céréales(q/h)	P (mm)
La rentabilité des céréales(q/h)	**Pearson** Correlation	1	**,896****
	Sig. (2-tailed)		**,000**
	N	10	10
P (mm)	**Pearson** Correlation	,896**	1
	Sig. (2-tailed)	,000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

فمن خلال الجدول السابق حصلنا على معامل الارتباط بقيمة 0,896 و أنها دالة في هذه الحالة عند مستوى 0,01 و ان N المشار إليها في الجدول هي عدد الحالات أو الأفراد و هنا لدينا N تساوي 10 و التي هي عشر مناطق فلاحية. و يمكن اختصار الجدول السابق إلى الجدول التالي :

Correlations Statistics=Pearson Correlation	La rentabilité des céréales (q/h)	P (mm)
La rentabilité des céréales (q/h)	1	0,896
P (mm)	0,896	1

نلاحظ من خلال الجدول أن الارتباط بين نفس المتغيرة هو دائما **1** و أن الارتباط بين المردودية و التساقطات هو نفسه بين التساقطات و المردودية و هذه قاعدة : أن الارتباط بين نفس المتغيرة يساوي **1** و

أن الارتباط بين X و Y هو نفسه الارتباط بين Y و X . و في هذه الحالة الارتباط بين التساقطات و مردودية الحبوب قوي جدا

و عند دراسة الانحدار توصلنا للجداول التالية :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,896[a]	,803	,778	2,198

a. Predictors: (Constant), P (mm)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	157,340	1	157,340	32,558	,000[a]
	Residual	38,660	8	4,833		
	Total	196,000	9			

a. Predictors: (Constant), P (mm)

b. Dependent Variable: La rentabilité des céréales(q/h)

يبين الجدول الأول قيمة معامل الارتباط R و معامل التحديد R^2 فالأولى هي نفسها التي حصلنا عليها سابقا و التي تساوي 0,896 أما معامل التحديد فيقدر ب **80**%

في حين أن الجدول الثاني يفسر لنا تحليل التباين ANOVA للانحدار و أن قيمة FO المحسوبة عند مستوى دلالة a = 0,05 و درجات حرية df1 = 1 و df2 = 8 تساوي 32,558 و هي أكبر من قيمة F1 الجدولية عند نفس المستوى التي تساوي 5.317 . و بناء عليه نرفض الفرضية الصفرية H0 و نقبل بالفرضية البديلة و نحكم بوجود علاقة خطية بين المتغير التابع و المتغير المستقل و بالتالي يمكننا الاستدلال على أن 80% من المردودية تفسرها كمية التساقطات بالمنطقة و أن 20% تعود لعوامل أخرى.

عند مستوى الدلالة 0,05 فإن الباحث يتأكد بنسبة **95**% من صحة فرضه و يخطئ فقط في **5**% من الحالات، ويجب تحديد مستوى الدلالة الذي سيعتمد عليه في اختبار صحة الفرض، و أهم مستويين هما **99**%، **95**%. أي (0,05 و 0,01)

إذن كيف يمكننا من خلال مستوى الدلالة تأكيد صحة الفرض أو نفيه؟

عند اجراء اختبار ما فإننا سنحصل على مستوى دلالة يرمز له بالرمز sig يتراوح معدله بين **0** و **100** أي بين (0,000 و 0,100) و يختلف حسب درجة أو مستوى الثقة التي نحددها، و غالبا ما نختار

0,05 كمستوى الثقة و هو المستوى المحدد افتراضيا من طرف برنامج SPSS أي أن الباحث سيقوم باختبار الفرضيات و سيحصل على نتائج تكون نسبة الصدق فيها ب 95% في حين أنه يمكن أن يخطأ في 5% و التي تساوي درجة الثقة التي حددها سابقا 5% أي (0,05).

يجب تحديد الفرضيات H0 (الفرضية الصفرية) و H1 (الفرضية البديلة) عند كل عملية اختبار كالارتباط و الانحدار أو مقارنة المتوسطات

فمثلا عند دراسة الارتباط تكون الفرضية الصفرية تقضي بعدم وجود ارتباط بين X وY

الفرضية الصفرية = لا يوجد ارتباط وثيق بين المتغير المستقل X و المتغير التابع H0 = Y

الفرضية البديلة = يوجد ارتباط ذو دلالة بين المتغير المستقل X و المتغير التابع H1 = Y

إذن عند إجراء الاختبار المتمثل في الارتباط يمكننا الاعتماد على درجة sig المحصل عليها من خلال برنامج SPSS و التي تكون كما يلي :

الفرضية	H_0	H_1
نتيجة sig	sig المحصل عليها أكبر من درجة الثقة المحددة أو أكبر من 0,05 المحددة افتراضيا	sig المحصل عليها أصغر من درجة الثقة المحددة أو أصغر من 0,05 المحددة افتراضيا
	0,05 < sig	sig < 0,05
القرار	قبول الفرضية الصفرية	رفض الفرضية الصفرية و قبول الفرضية البديلة

إذا كانت sig المحصل عليها أكبر من sig المفترضة (درجة الثقة) نقبل الفرضية الصفرية H0 التي تقضي بعدم وجود دلالة أو عدم وجود ارتباط.

أما إذا كانت sig المحصل عليها أصغر من sig المفترضة (درجة الثقة) فإننا نرفض الفرضية الصفرية H0 و نقبل بالفرضية البديلة H1 التي مفادها وجود دلالة احصائية أو وجود ارتباط بين X و Y

في حالة مقارنة المتوسطات فإن الفرضية H0 الصفرية يكون مفادها تساوي المتوسطات و أن الفرضية البديلة H1 تقضي بوجود فوارق أو اختلافات ذات دلالة احصائية و تعود للفئة ذات المتوسط الأكبر في حالة اختبار T.

ترجمة بعض المصطلحات الخاصة ببرنامج SPSS

English	Francais	عربي
Correlations	Corrélations	الارتباط
Cumulative Percent	Pourcentage cumulé	نسبة التكرارات التراكمية
df	ddl	درجات الحرية
Frequency	Effectifs	التكرارات
Maximum	Maximum	أعلى قيمة
Mean	Moyenne	المتوسط
Mean Square	Moyenne des carrés	متوسط المربعات
Median	Médiane	الوسيط
Minimum	Minimum	أدنى قيمة
Missing	Manquante	الحالات أو القيم المفقودة
Mode	Mode	المنوال
N	N	عدد الأفراد أو الحالات
Percent	Pourcentage	النسبة
Percentiles	Centiles	الأرباع أو المئينات
Range	Intervalle	المدى
Regression	Régression	الانحدار
Residual	Résidu	الباقي
Sig (significant)	Sig (significative)	مستوى الدلالة
Std. Deviation	Ecart-type	الانحراف المعياري
Sum	Somme	المجموع
Sum of squares	Somme des carrés	مجموع المربعات
Valid	Valide	الحالات المحسوبة
Valid Percent	Pourcentage valide	نسبة القيم المحسوبة فقط
Variance	Variance	التباين (التغير)
2-tailed	bilatérale	في اتجاهين

القيم الجدولية لاختبار F عند مستوى الدلالة a = 0,05

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

معادلات جغرافية

** تقرأ جميع المعادلات من اليمين إلى اليسار حسب اتجاه الكتابة.

معدل الخصوبة :
$$\frac{\text{عدد المواليد الأحياء في السنة}}{\text{عدد النساء في سن الإنجاب}(15-49)} \times 1000$$

معدل وفيات الاطفال الرضع :
$$\frac{\text{عدد وفيات الأطفال الأقل من سنة}}{\text{عدد المواليد الأحياء خلال نفس السنة}} \times 1000$$

معدل الوفيات الخام :
$$\frac{\text{عدد الوفيات في السنة}}{\text{عدد السكان الكلي في منتصف السنة}} \times 1000$$

معدل الزيادة الطبيعية :
$$\frac{\text{معدل المواليد الخام} - \text{معدل الوفيات الخام}}{\text{عدد السكان في منتصف السنة}} \times 1000$$

نسبة الإعالة :
$$\frac{\text{السكان غير القادرين على العمل}}{\text{السكان القادرون على العمل}} \times 100$$

متوسط نصيب الفرد من الدخل القومي :
$$\frac{\text{إجمالي الدخل القومي بالدرهم}}{\text{عدد السكان}} = \text{درهم لكل فرد}$$

إنتاجية الأراضي الزراعية :
$$\frac{\text{حجم الإنتاج}}{\text{مساحة الأراضي الزراعية}}$$

الكثافة السكانية :
$$\frac{\text{عدد السكان}}{\text{المساحة (كلم}^2\text{)}} = \text{نسمة\كلم}^2$$

درجة التزاحم :
$$\frac{\text{عدد السكان}}{\text{عدد الغرف}} = \text{شخص\غرفة}$$

كثافة الغرف :
$$\frac{\text{عدد الغرف}}{\text{المساحة المبنية (كلم}^2\text{)}} = \text{غرفة\كلم}^2$$

نسبة التحضر :
$$\frac{\text{عدد السكان الحضريون}}{\text{عدد السكان الإجمالي}}$$

متوسط حجم الأسر : $\frac{\text{عدد السكان}}{\text{عدد الأسر}}$

معدل الهجرة الوافدة : $\frac{\text{عدد المهاجرين إلى المنطقة}}{\text{عدد سكان المنطقة}} \times 100$

معدل الهجرة النازحة : $\frac{\text{عدد المهاجرين من المنطقة}}{\text{عدد سكان المنطقة}} \times 100$

معدل الهجرة الصافية : $\frac{\text{عدد الوافدين} - \text{عدد النازحين}}{\text{عدد السكان}} \times 100$

معدل الهجرة الكلية : $\frac{\text{عدد الوافدين} + \text{عدد النازحين}}{\text{عدد السكان}} \times 100$

مؤشر الانعطاف : $\frac{\text{طول الطريق الفعلي}}{\text{طول الطريق بخط مستقيم}} \times 100$

الكثافة السياحية : $\frac{\text{عدد الليالي السياحية}}{\text{عدد السكان}}$

طاقة الاستيعاب : $\frac{\text{مجموع الأسرّة}}{\text{متوسط مدة الإقامة}} \times 365$

متوسط التدفق السياحي : $\frac{\text{عدد السياح}}{\text{عدد السكان}} \times 1000$

متوسط مدة الإقامة السياحية : $\frac{\text{عدد الليالي السياحية}}{\text{عدد السياح}}$

لا تبخلوا من فضلكم بدعوة صالحة في حق من سهرا على تربيتي (والدي الكريمين).
و أتمنى أن يكون هذا العمل صدقة جارية في حقهما.